Xrootd and Distributed Storage

OSG Site Admin's Workshop
8/9/2011

What is Xrootd

Developed by SLAC, CERN

- http://xrootd.slac.stanford.edu/

A file access and data transfer protocol using a distributed architecture.

Defines POSIX-style byte-level random access for

- Arbitrary data organized as files
- Identified by a hierarchical directory namespace

Xrootd is the reference implementation of this protocol.

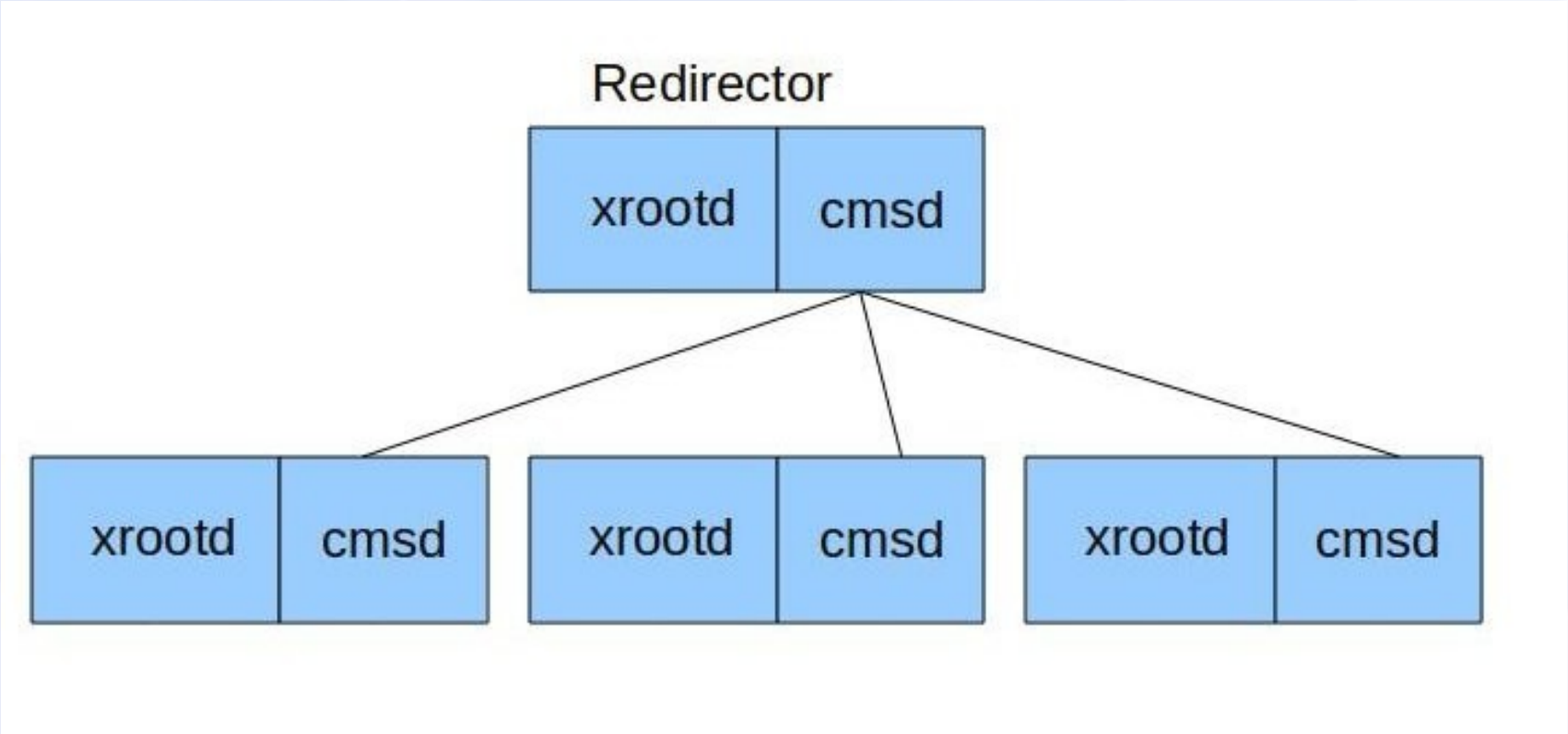
How to Access Xrootd

- Root framework used by HEP experiments
- POSIX FUSE access through XrootdFS
- POSIX pre-load library for full POSIX access.
- xrdcp xrootd copy utilities
- SRM access through BeStMan
  - BeStMan uses FUSE directly or can use GridFTP which uses POSIX Pre-load

Xrootd Features

- Clusters highly disparate systems
- Can use a variety of underlying storage systems
- Hierarchy of redirectors can scale system exponentially.

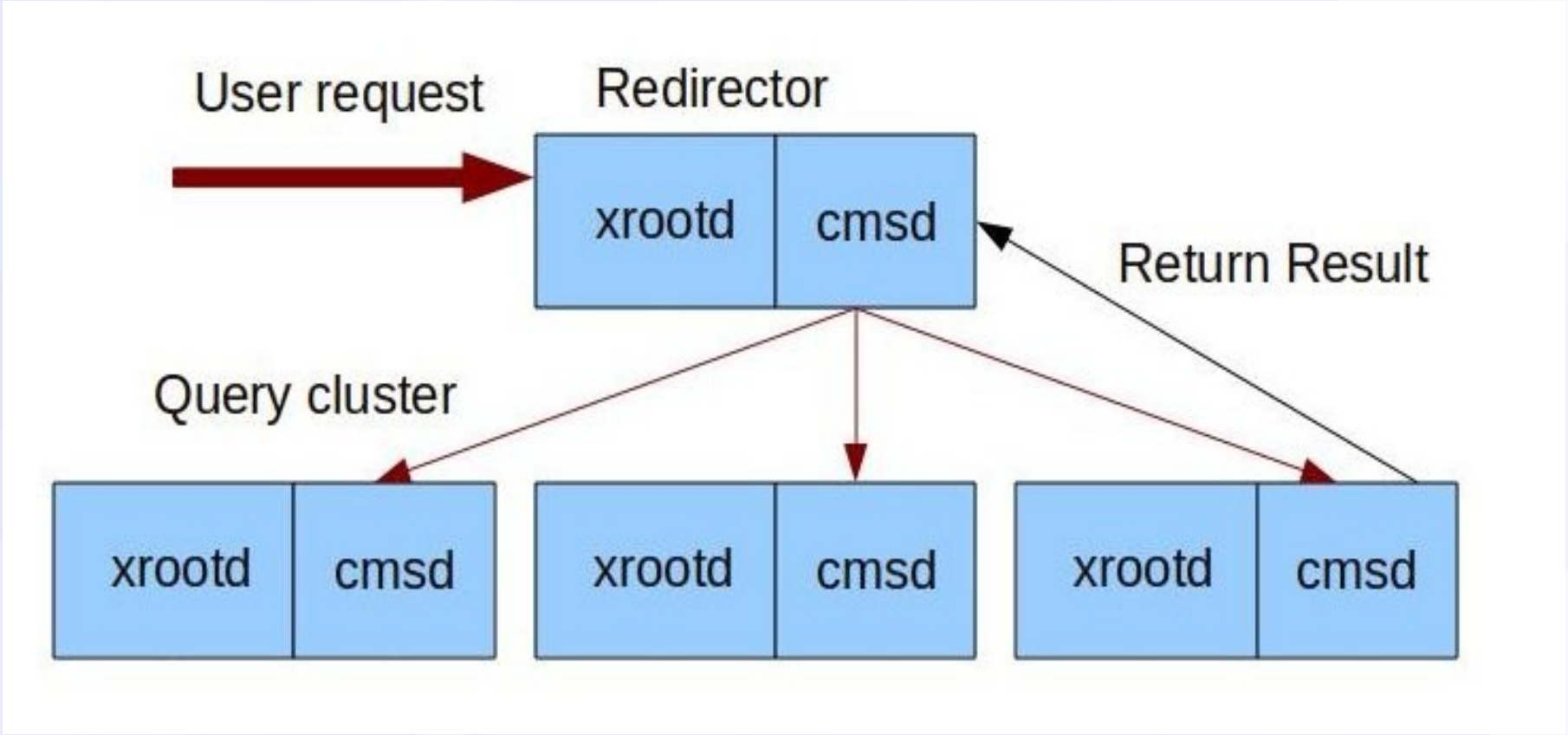
Xrootd Architecture



xrootd: daemon to manage storage

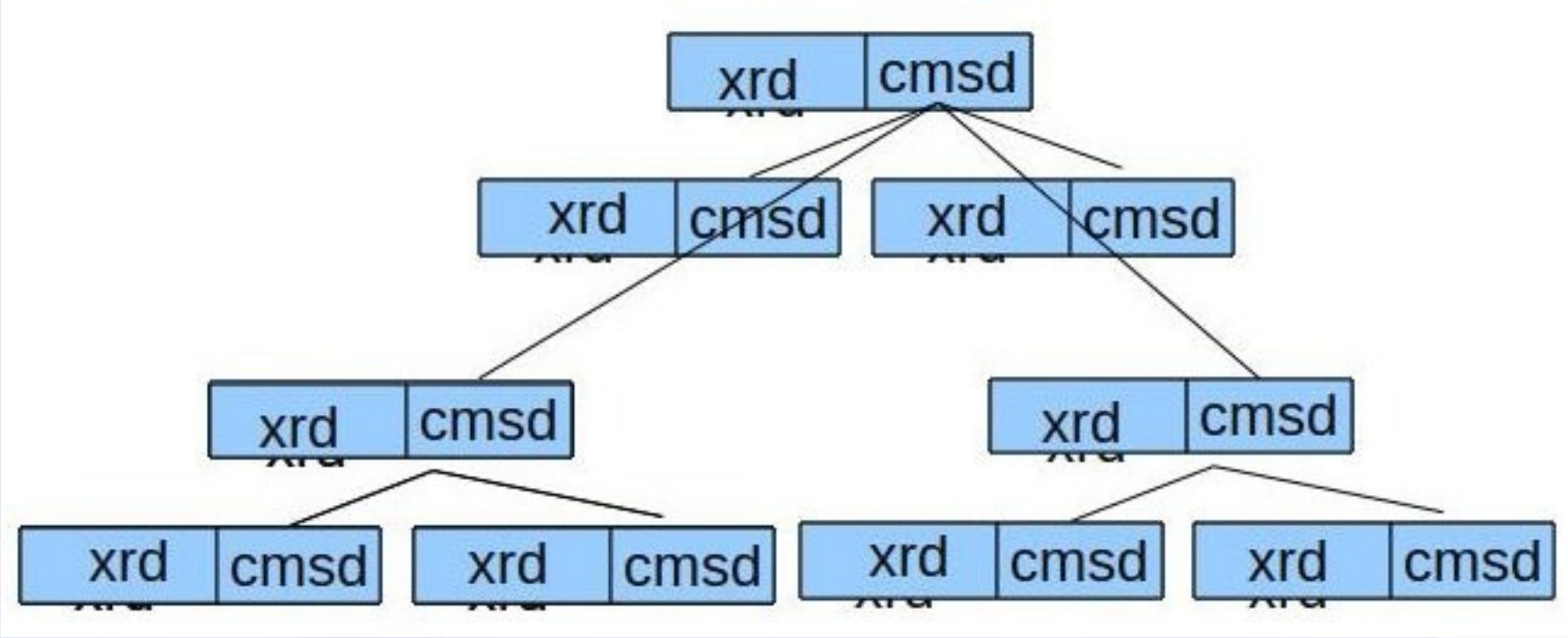
cmsd: daemon to manage cluster

Xrootd Clusters



- User queries xrootd (redirector)
- Redirector queries all xrootd in cluster

Xrootd Clusters



- Real power of xrootd is that clusters can be combined into a global hierarchy
- Queries to global redirector query local rdrs

Xrootd Clusters - FRM

- FRM = file residency manager
- Requests can come to a local re-director or to a regional/global redirector.
- Sites cache files and use local files when possible. When a file is not available,
  - 1) Redirectors query all subsidiary re-directors. Take first available.
  - 2) If no re-directors respond, then the query is forced up to a more global re-director.
  - 3) File transfer from remote to local node

Federated Storage

- Sites are joined in a **common namespace**
- Each site can modify its own (access) rules
- Scalability increases as more sites join
  - Overhead only increases logarithmically
- Can copy across different architectures
- Can copy across different administrative domains (plugins for firewall exist)

Xrootd Plugins

- Xrootd is very “pluggable”
- Plugins for:
  - Authentication (krb, ssh, gsi)
  - Authorization (dbms, voms)
  - Protocol driver (xrd)
  - Logical file system (ofs, sfs, alice, etc)
  - Physical file system (ufs, hdfs, hpss, etc)
  - Prefix encoding (lfn2pfn)

Xrootd with CMS

- Global Redirector
- Regional Redirector (EU & USA)
- Each site has its own redirector

- CMSSW uses Xrootd as a fallback option
  - If files are missing

Xrootd with CMS

- Currently using xrootd: T2_US_Nebraska, T2_US_Caltech, T2_US_UCSD, T3_US_FNALLPC
- Sites use various levels of caching from completely diskless (Omaha) to full Xrootd install (T3_US_UCR)
- Xrootd has been integrated with dCache at FNAL tier 1 to provide root-based access to dCache data.

Xrootd with Atlas

- Goal: Any data, any time, any where
- Before xrootd: If program was missing data (dataset was moved/deleted/broken), program immediately FAILED.
- Intermediate: If site does not have a data set, try xrootd to get it. If it can transfer it, it will continue on.
- Eventually: Data will not need to be staged, and everything will be xrootd cache-driven.

Xrootd Demonstrator

- Built on top of File Residency Manager (FRM)
  - Controls residency of files
  - Locally configured to handle events:
    - A requested file is missing
    - A file is created or an existing file is modified
    - Disk space is getting full
  - Grabs files from redirector “when necessary”
  - Policy different for ATLAS vs ALICE

For more information

- Xrootd
  - http://xrootd.slac.stanford.edu/
- OSG – RPM Installation coming soon
  - Documentation still in progress:
  - https://twiki.grid.iu.edu/bin/view/SoftwareTeam/XrootdRPMPhase1
- CMS Xrootd
  https://twiki.cern.ch/twiki/bin/view/Main/CmsXrootdArchitecture